THE SPIDER, STUCK IN A WEB

Robert Filman • Lockheed Martin Advanced Technology Center • filman@stc.lockheed.com
Feniosky Peña-Mora • Massachusetts Institute of Technology • feniosky@mit.edu

MEASURING INTERNET CONGESTION

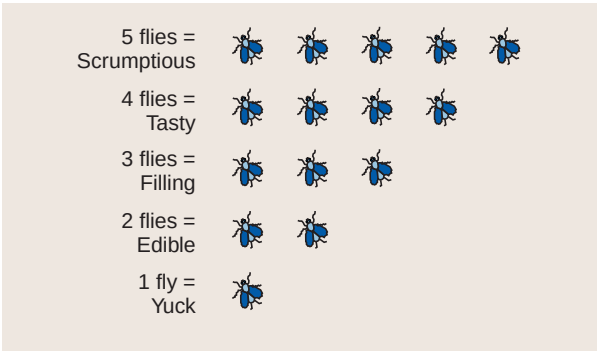
http://nic.merit.edu/ipma
Visited mid February 1997

Internet traffic is often compared to weather. To talk about the weather, you have to be able to characterize it. The Internet Performance Measurement and Analysis project is an NSF-sponsored research project of the University of Michigan and Merit Network to develop tools to characterize the Internet weather (traffic) and to actually measure it. This site asserts that "router instability"—the rapid fluctuation of Internet routing information—appears to be growing at a faster rate than the Internet itself, despite the development of more stable routing algorithms. This claim is roughly equivalent to warnings of global warming.

"Everyone complains about the weather, but no one does anything about it." The press page (http://nic.merit.edu/ipma/press/) at this site has an entertaining collection of (primarily press) clippings predicting the Internet's failure and a perspicacious set of rebuttals from engineers, suggesting that because Internet weather is a function of algorithms and economics, we may in fact be able to do something about it. The site also includes a fair collection of graphs illustrating current net congestion.

Bob rates the site: 
I found the contrasting views of network deterioration (between the net engineers, the press, and—somewhere in-between—the reigning Net Cassandra, Bob Metcalfe) entertaining. I'm sure the graphics were illustrating something, though I'm not always sure what. They took so long to download, I couldn't really check them out.

Feniosky rates the site: 
If you are really interested in knowing the current status of the Internet in terms of its life expectancy and structural conditions, then this site will help you get either a clearer or more confused picture of the situation. The site provides information about Internet instability in terms of announcements, withdrawals, and average number of routing table entries. It also presents data in both table and graph formats, and provides links to the opinions of "Internet engineers" on its future. There is a lot information here. After perusing it, you may conclude that the Internet can't possibly survive its own popularity or that smart people will find a way to avoid its collapse. In either case, this site is the result of a project looking to provide data that will help us understand the situation and perhaps find a way to do something about it.



http://www.mids.org
Visited mid February 1997

What's a weather report without a weather map? Matrix Information and Directory Services produces maps of Internet weather, using colors and circles and movies to show where the traffic is.

Bob rates the site: 
Well, perhaps even movies, except the movies didn't work on my system. Very pretty (and how often do I see a US map with Mountain View but not San Francisco?) but not quite as useful as the TV news.

Internet weather seems too ephemeral for forecasting. After all, if there's a cold front moving in from Canada, then pretty soon it's going to get cold. A lot of network activity in Mountain View right now is not nearly as predictive of the network activity in San Jose a few hours later. I give the site points for style, though.

Feniosky rates the site: 
This site is totally confusing and has almost everything you can imagine. It doesn't just give you the weather (by the way, I don't know why they call it weather if what they are measuring is traffic; so it should be traffic forecast), but it also gives you links to places where you can find about your genealogy. The information on the Web is old. For example, information about the number of users of the Internet is from October 1995. The graphs are so small that you can't see them. If you want bigger graphs you have to subscribe to their service.

This site may have a lot useful information but I definitely did not see it, and I am not going to subscribe to their service to find out if it gets any better.

EASING INTERNET CONGESTION

http://www.nlanr.net/Cache/
Visited mid February 1997

The National Laboratory for Applied Network Research is an NSF-funded, collaborative project of American supercomputing centers for studying and evolving the American network infrastructure. This site describes the NLANR cache prototype project.

While it's straightforward to make a Web client remember pages it has already seen, client caching can only make a minor dent in overall Net traffic. Many people may want to read the cover article in today's *Wall Street Journal* or the latest CNN on the OJ case, but only a few will want to read that article more than once. Because such pages are popular, they are repeatedly redistributed over long network routes. By storing copies of popular pages at critical intersections around the net, it is possible to reduce Internet traffic considerably. To quote the site, "A 1993 study of FTP traffic on the NSFnet backbone concluded that several well-placed caches could reduce FTP traffic by 44 percent." The NLANR Cache project implements such a cache (in a hierarchical form) and measures the net effect.

The site also includes links to pages about network traffic, showing graphs of relative activity and a page (http://www.nlanr.net/NA/) that provides a wide-ranging insight into current Internet traffic loads. The authors argue that Internet congestion can be ameliorated by (1) using multicast for multicast audio and video applications (though non-multicast video like cu-seeme is a major threat), (2) using future versions of HTTP that coalesce multiple small transactions into single requests, (3) caching popular documents intelligently throughout the Net, (4) using information discovery tools that develop information brokers, rather than simple web-crawlers, (5) finding better pricing mechanisms, and (6) integrating security into Net mechanisms without sacrificing performance.

Bob rates the site: 
While I found the explanation of the actual cache project to be weak, there's a wealth of other useful information in the surrounding pages. I learned that Web traffic is bad in the morning, takes a break at noon for lunch, is worst in the afternoon, and doesn't really ease up until nine or ten in the evening. Web traffic and multimedia broadcast are both major consumers of network bandwidth, but video teleconferencing (like cu-seeme), with both large bandwidth demands and only a single user destination, is more threatening still. On the other hand, the site suffers from the "maze-of-twisty-passages" effect—-there are too many links to other things, without an overall sense of topography.

Feniosky rates the site: 
The concept of sharing resources and creating a hierarchy of resource usage is really appealing even though it is not new. The idea of easing congestion of popular sites by caching the information in certain strategically located servers helps redistribute the traffic load to a site. People in New Zealand might access Web pages from United Kingdom faster in a cache server in the United States, if the UK site is congested. Documentation presented in some of the links on this site show the idea of cache servers being developed in several places concurrently. The site is informative and proposes a different way to decongest the Net.

http://harvest.transarc.com/
Visited mid February 1997

Harvest was an ARPA-funded project, centered at the University of Colorado, Boulder, to develop an integrated toolset for gathering, extracting, organizing, searching, caching and replicating information on the Net. Typical Net "robots" search by successively calling up pages and their links—a process that is expensive in Net resources. The Harvest project observes that (1) indexers that share information spare the Net the cost of new searches; (2) information is often stored in particular formats, and indexing can be more efficient if the searcher can easily describe that format; and (3) indexing systems are themselves popular, which leads to such sites becoming bottlenecks.

Harvest provides tools for building specialized indexers and caching, sharing, and replication mechanisms to ease and decentralize the searching load. Examples of indexes built with Harvest (and nominally reachable from this page) include WWW homepages, PC software, State laws, the CIA, and Library of Congress. The site argues that these indexes are stored in a more semantically accessible form than simple names.

Bob rates the site: 
I enjoyed the quality of explanation of this site (and the ability to download papers), but the actual behavior of the system over links and searches was less impressive. The WWW homepages and PC software search sites produced "file not found" errors, and the State laws search revealed no matches to "sex," "taxes," "tax," or "death" (really!). On the other hand, the CIA index responded to a search for "Bosnia" with five pages of maps and data, including links to places like "The Former Yugoslav Republic of Macedonia" (which presumably is now only known by some icon, a further problem for search engines).

Feniosky rates the site: 

This site provides information about the software used in the NLANR cache prototype project reviewed earlier. The documentation is presented well, and the explanation here actually makes it easier to understand the NLANR project. I found myself going back and forth between the two sites to get a good grasp of what was going on and what was differ-

ent. To tell the truth, the information in these two sites is so intertwined that I would not encourage reading one without the other. The site is a "success" story of university research becoming commercial.

There is another site at http://www.netcache.com/, which describes the commercial version of Harvest. I find the research site's documentation much less confusing than the commercial site's, but one interesting thing in the commercial site is a paper, "A Hierarchical Internet Object Cache," published on the 1996 Usenix technical conference. A section on deployment says "Deploying any Web cache—Harvest, Netscape, or CERN—for a regional network or Internet Service Provider is tricky business. As Web providers customize their pages for particular browsers, maintaining a high cache hit rate will become harder and harder." Are they crying for standards or telling us "buy at your own risk"?

CAUSING INTERNET CONGESTION

http://pioneer.pointcast.com/tour/
Visited mid February 1997

Everyone complains about the weather, but how many of us can brag about doing something that makes it noticeably worse? Pointcast promises to deliver to your desktop news that is customized to your interests (along with the advertisements that support the service), perhaps as a screen-saver so the information can get there just at the moment you're not looking at the computer.

Bob rates the site: 

I won't comment on the quality of the Pointcast service itself, but the Net site itself was a disappointment. I got a simple sequential tour, with pasted-in images. I expected a more dynamic demonstration, with actual information flying at me through the ether, most likely prompted by my particular enthusiasms. Maybe they didn't want to overload their server.

Feniosky rates the site: 

I have Pointcast on my computer. I receive their information. I am a sinner. Well, I am a reformed sinner. At the beginning, I selected almost all the channels that they offer. Then I realized it was too much information and that I never looked at it. So now I mostly use it as an interesting screensaver.

Well, going back to the Web site: I always found the contrast between their site and their screen mind-boggling. The screen is so active that it almost looks like a TV with advertising, action, hype, and never-ending dynamism. The Web site is all the contrary: very passive, sequential, and boring.

We have a saying in Spanish: "In the house of a blacksmith, wooden knife."

PROGRAMMING TOOLS

http://www.infospheres.caltech.edu/
Visited mid February 1997

The Caltech Infospheres Project is developing theory and tools for persistent, concurrent systems composed of distributed objects. The site introduces the project, links to a number of papers, and lets you download the current version of their Java middleware.

Bob rates the site: 

I found this site to be too long on generalities and too short on specifics. It took me a while to get a feeling for what this group is doing, and even after reading a couple of the papers, I don't think I could pass a test on what's special about it. More to the point, for a site devoted to creating distributed Java applications, I hoped for a snazzy demonstration of a composed, collected computer consciousness, but I could find nary a hint of a running code—not even a bad demo.

Feniosky rates the site: 

I found the description and accompanying papers interesting. I found the site well documented and very well organized. It includes papers to read and code to download. The site is the storefront of a research center, and they did a good job in presenting their research.

The only problem I had with the site was the incompatibility of the title and their warning. Their title says "Bringing Theory to the Marketplace," then their warning says "Please Note! The goal of our group is to develop theories, algorithms, and prototype software to support a worldwide object network. This code is the alpha release of a part of the prototype system that we are using to explore ideas and example applications. It is not meant to be a competitor to CORBA, DCOM, or similar systems. It is a flexible, lightweight distributed object framework written in Java. We expect to add functionality to the prototype system over the next year." They talk of the marketplace but admit that their stuff is not ready to compete. Again, this site is interesting as a research site but perhaps oversold in its title. I was expecting more than papers. I was expecting demos that will show me what they so beautifully talk about, but what I got was a software to be downloaded (which I did not download).

ABOUT THE TOURIST

The Arachnoid Tourist scours the Net to find and review Web sites of interest to our readers.

What makes a site interesting? The Tourist appreciates style but cares most about content. Each issue we visit five to 10 sites and report on what we find and how well it works.

We welcome your suggestions for places to visit.